Evolutionary History of Gene Regulation in Bacteria

Morgan N. Price[1,2], Paramvir S. Dehal[1,2*], **Adam P. Arkin**[1,2,3]

[1]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov/; [2]Lawrence Berkeley National Laboratory, Berkeley, CA, 94720; [3]Department of Bioengineering, University of California, Berkeley, CA 94720

Analysis of gene regulation in ESPP bacteria relies on comparisons to model organisms, and hence on assumptions about how gene regulation evolves. To test these assumptions we examined the evolutionary histories of transcription factors and of regulatory interactions from the model bacterium Escherichia coli K12. We show that although most transcription factors have paralogs, these usually arose by horizontal gene transfer rather than by duplication within the E. coli lineage, as previously believed. Most neighbor regulators -- regulators that are adjacent to genes that they regulate -- were acquired by horizontal gene transfer, while most global regulators evolved vertically within the gamma-Proteobacteria. Neighbor regulators are often acquired together with the operon that they regulate, which suggests that the proximity is maintained by repeated transfers, and also aids the prediction of the regulators' function. Because of the complex evolutionary histories of most transcription factors, bidirectional best hits tend to be misleading, and most annotations of bacterial regulators are probably incorrect.

When we analyzed the histories of regulatory interactions, we found that the evolution of regulation by duplication was rare, and surprisingly, many of the regulatory interactions that are shared between paralogs result from convergent evolution. Furthermore, horizontally transferred genes are more likely than other genes to be regulated by multiple regulators, and most of this complex regulation probably evolved after the transfer. Finally, gene regulation is often not conserved, even within the gamma-Proteobacteria. Our results suggest that the bacterial regulatory network is evolving rapidly under positive selection. Such rapid rewiring of gene regulation may be crucial for adaptation to new niches.

Rapid Evolution of Gene Regulation

- Transcription factors (TF) have complex histories
  - Rampant transfer (not duplication)
    - "Orthologs" are problematic
    - Annotations are usually incorrect
  - "Neighbor regulation"
    - driven by horizontal gene transfer, aids annotation
  - Global regulators are more conserved
    - but we can predict little about other divisions
- Regulatory interactions not highly conserved

Vertical Inheritance of Global Regulators in *E. coli*

Gene tree for crp
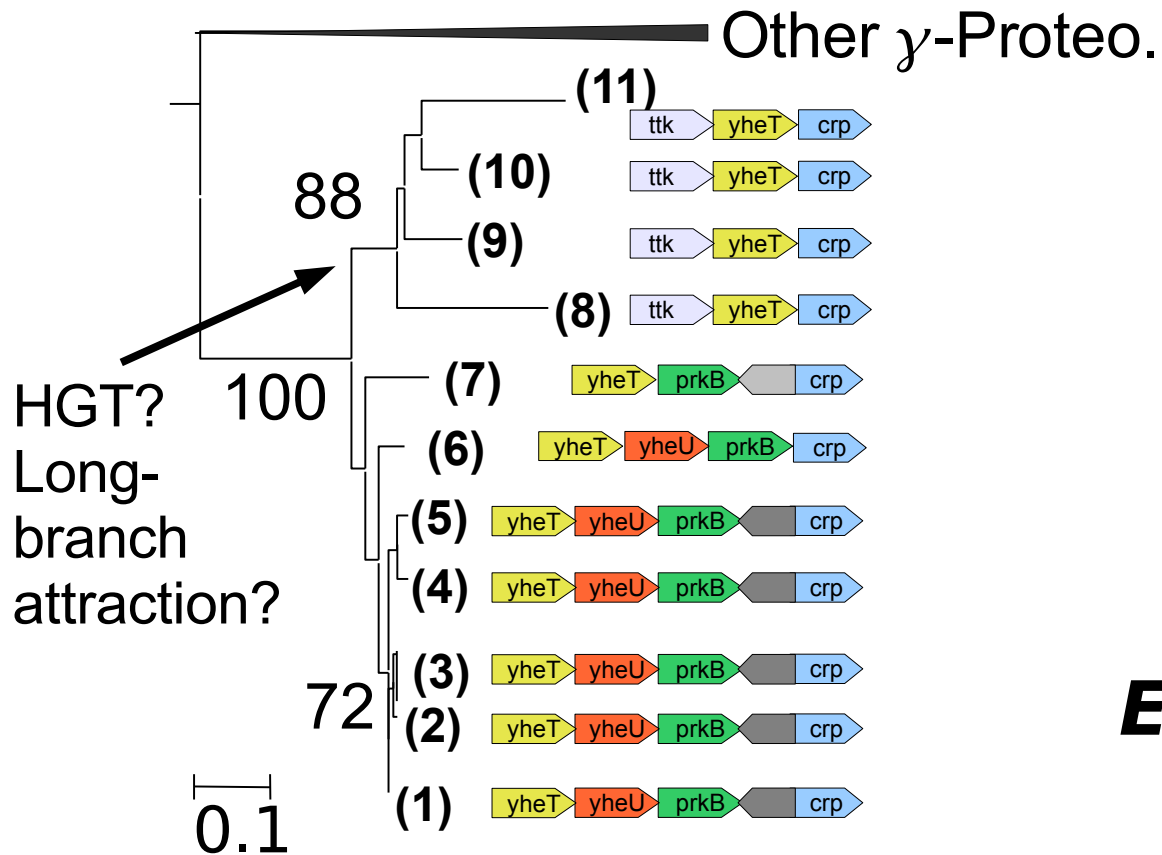
Distribution of crp
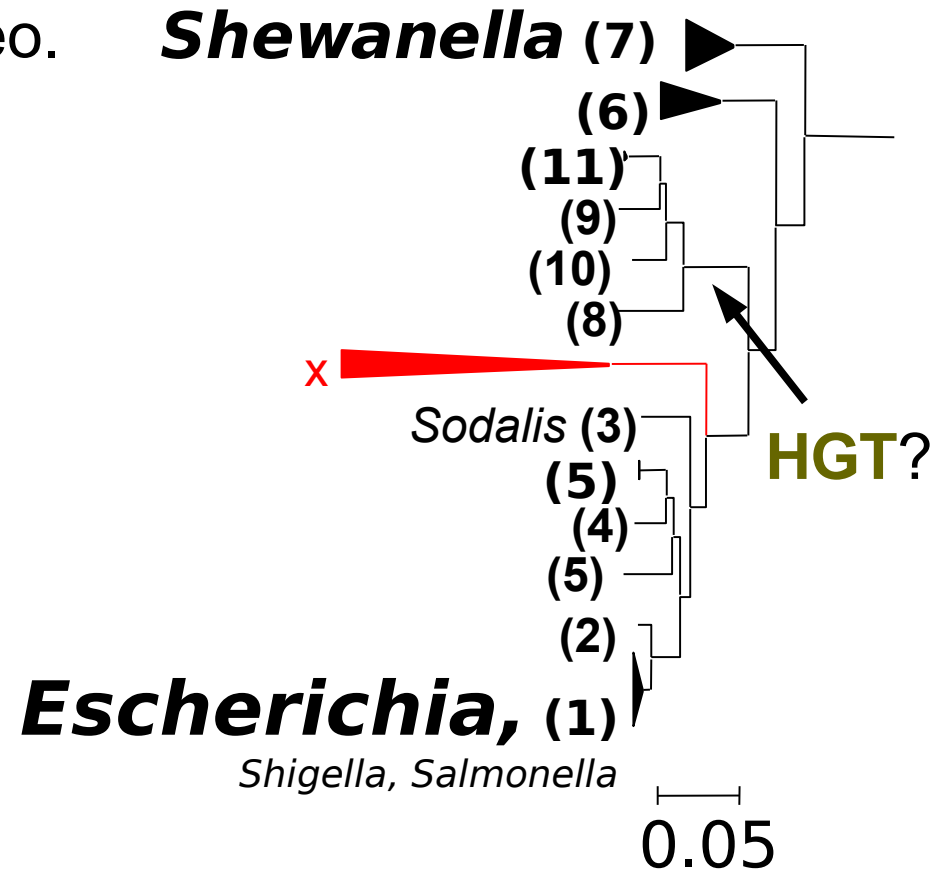
17/20 top global regulators are native

Complex Histories
of Neighbor Regulators in *E. coli*

Gene tree for xapR

Diverse Proteobacteria
No xapA or xapR nearby

xapR — xapA **(6)**

xapA | xapR **(5)**

xapA | xapR **(4)** ↰ HGT??

xapA | xapR **(3)**

xapA > xapB > xapR **(2)**

xapA > xapB > xapR **(1)** *E. coli*

98

80

0.2

Distribution of xapR

Shewanella **(4)**

HGT or
9 losses

Vibrio **(5)**
Vibrio **(3)**

(6)

HGT or
3 losses

(2)

(1)

0.05
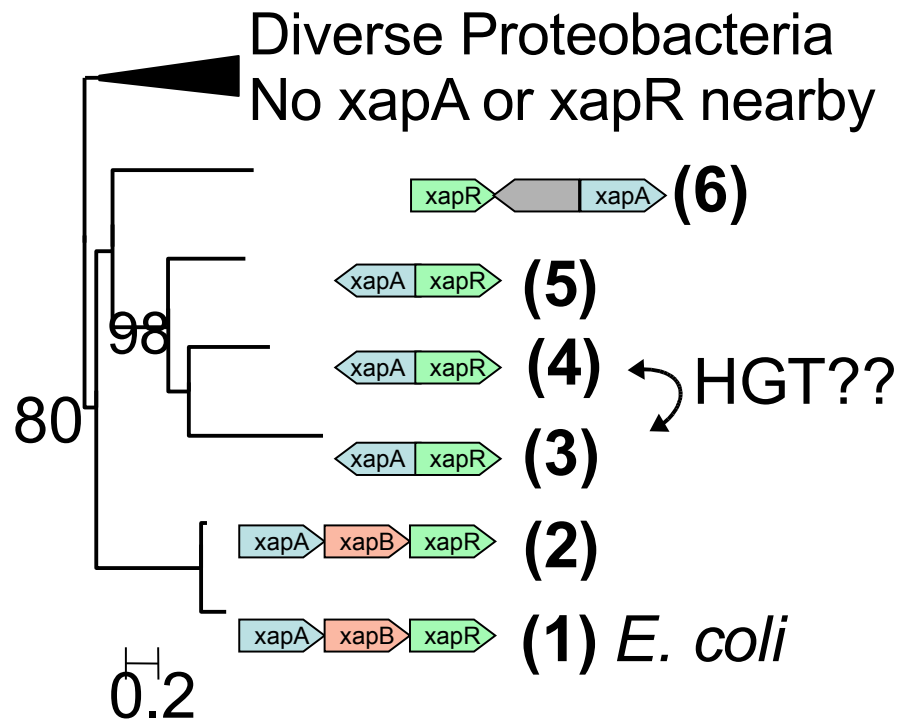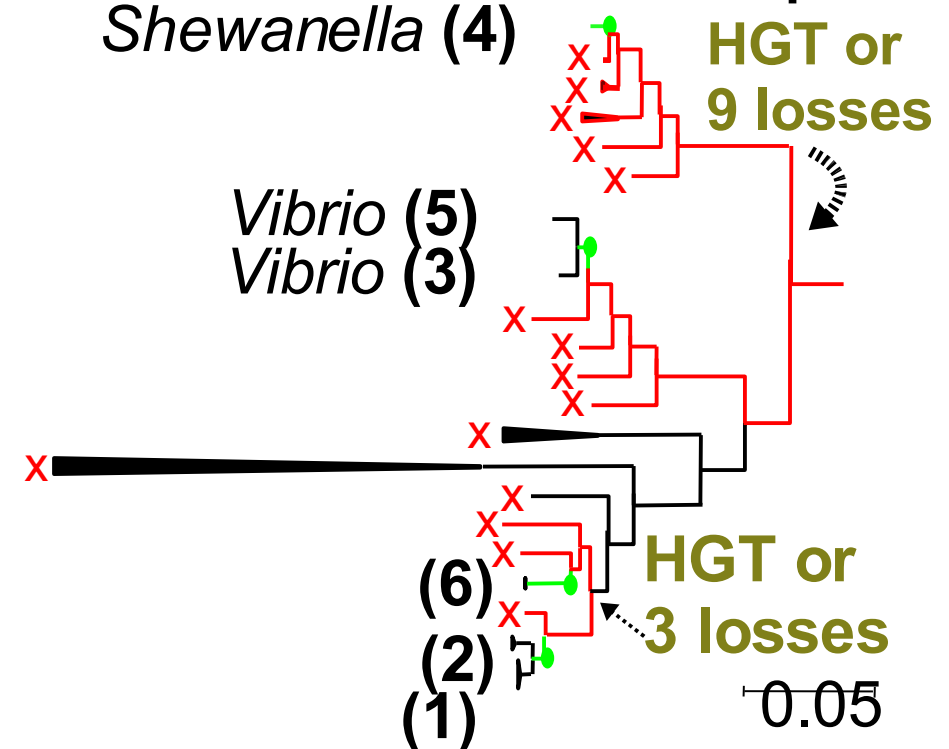
- Co-transfer: ~60% of neighbor regulators, 45% of putative regulators => predictions

- Repeated HGT: ~40% of neighbor reg.

Duplication of rbsR/purR

Gene tree for rbsR, purR

Firmicutes

β-Proteo

100

99

dup.

77

0.2

rbsD rbsA rbsC rbsB rbsK **rbsR** **(13)**

rbsD rbsA rbsC rbsB rbsK **rbsR** **(12)**

rbsD rbsA rbsC rbsB rbsK **rbsR** **(11)**

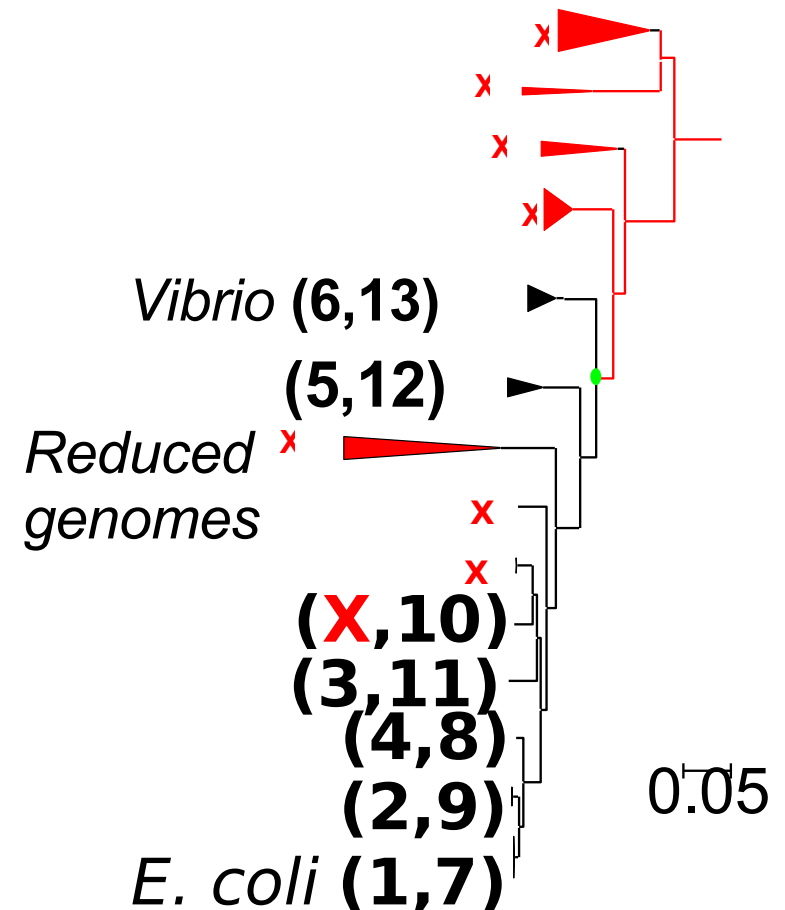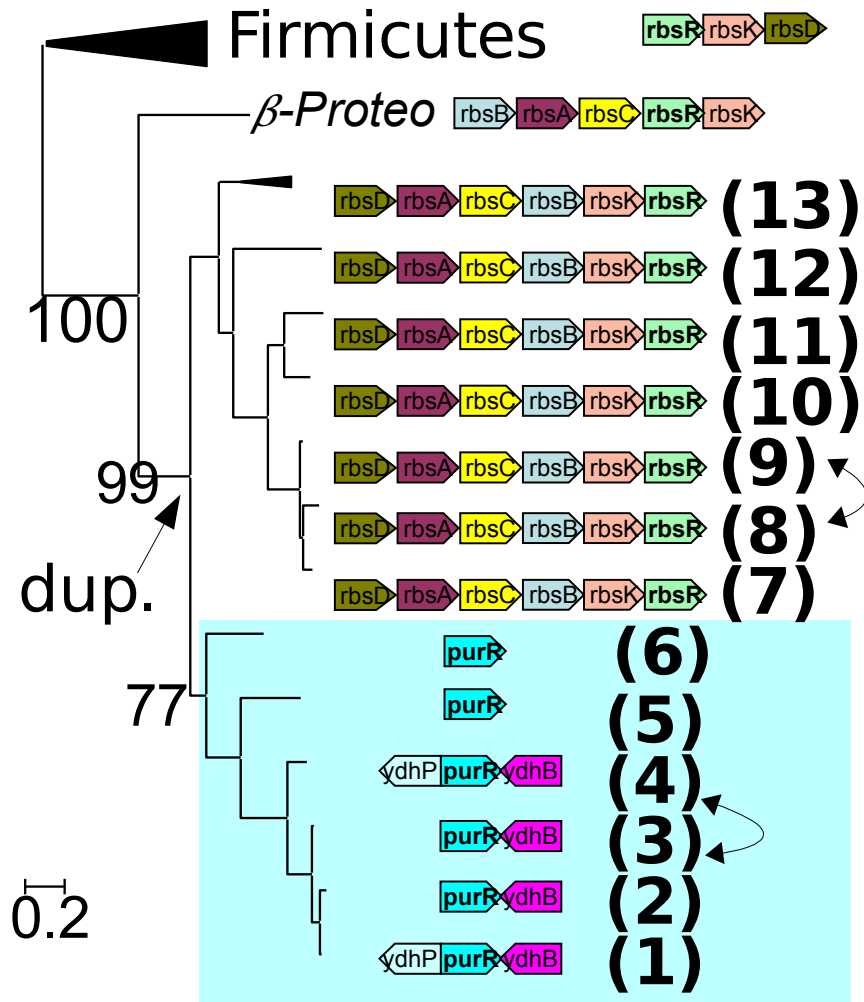rbsD rbsA rbsC rbsB rbsK **rbsR** **(10)**

rbsD rbsA rbsC rbsB rbsK **rbsR** **(9)**

rbsD rbsA rbsC rbsB rbsK **rbsR** **(8)**

rbsD rbsA rbsC rbsB rbsK **rbsR** **(7)**

purR **(6)**

purR **(5)**

ydhP **purR** ydhB **(4)**

purR ydhB **(3)**

purR ydhB **(2)**

ydhP **purR** ydhB **(1)**

rbsR rbsK rbsD

rbsB rbsA rbsC **rbsR** rbsK

Distribution of rbsR, purR

Vibrio **(6,13)**

(5,12)

Reduced genomes

(X,10)

(3,11)

(4,8)

(2,9)

E. coli **(1,7)**

0.05
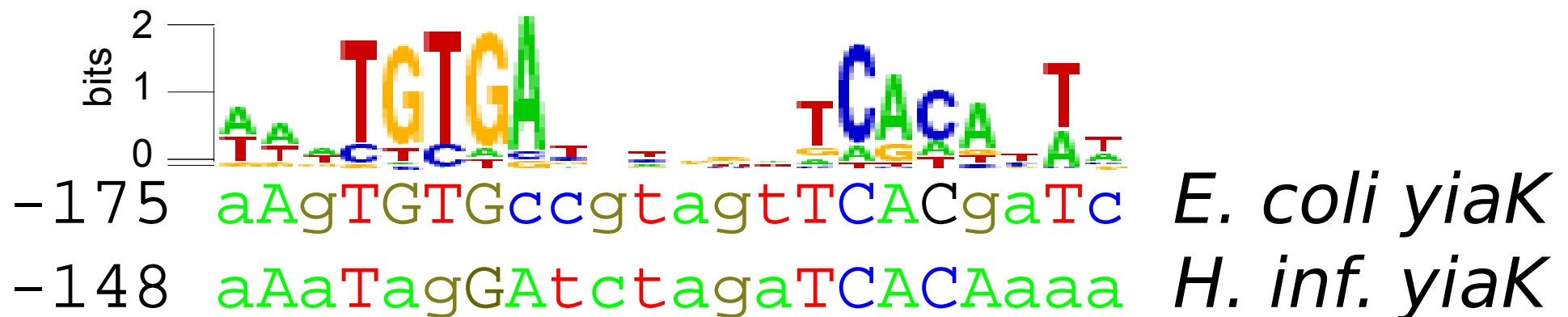
- Dups are rare (13% of *E. coli* TFs)
- Non-overlapping functions (~half of dups)
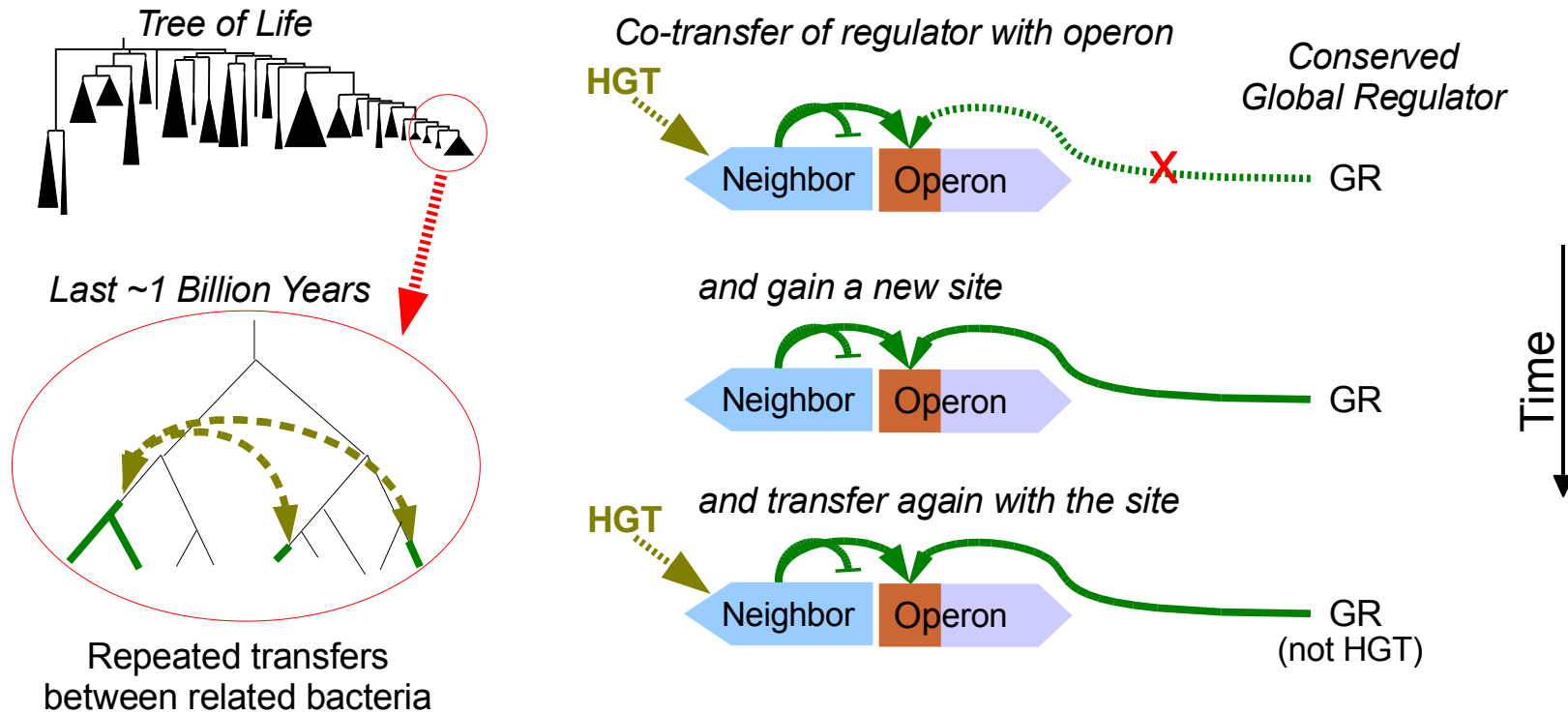
Sources of the CRP Regulon

- From distant bacteria w/o CRP (~80%)

- From related bacteria with CRP (~20%)
  - CRP site conserved across HGT in 4/12

- Sites usually not conserved across HGT (6/20 for global regulators)
  - except for co-transfer (presumably)

-175 aAgTGTGccgtagtTCACgaTc *E. coli yiaK*

-148 aAaTagGAtctagaTCACAaaa *H. inf. yiaK*

Apparent Evolution of Regulation by Duplication is Convergent

Regulation:

dcuR → dctA

arcA ↗

dcuR from Firmicutes,
dctA from distant γ-Proteo.

acquire arcA (or
duplication from
torR)

	dcuR	arcA	dctA
Escherichia etc.	+	+	+
Salmonella	+	+	+
Klebsiella	+	+	+
Photorhabdus	-	+	+
Erwinia	+	+	+
Yersinia	-	+	+
Sodalis	-	+	-
	-	-	-
Pasteurellaceae	-	+	-
Photobacterium	-	+	-
Vibrio	-	+	-
Shewanella	-	+	-
Colwellia, ...	-	+	-
Acinetobacter, ...	-	-	+
Pseudomonas, ...	-	-	+
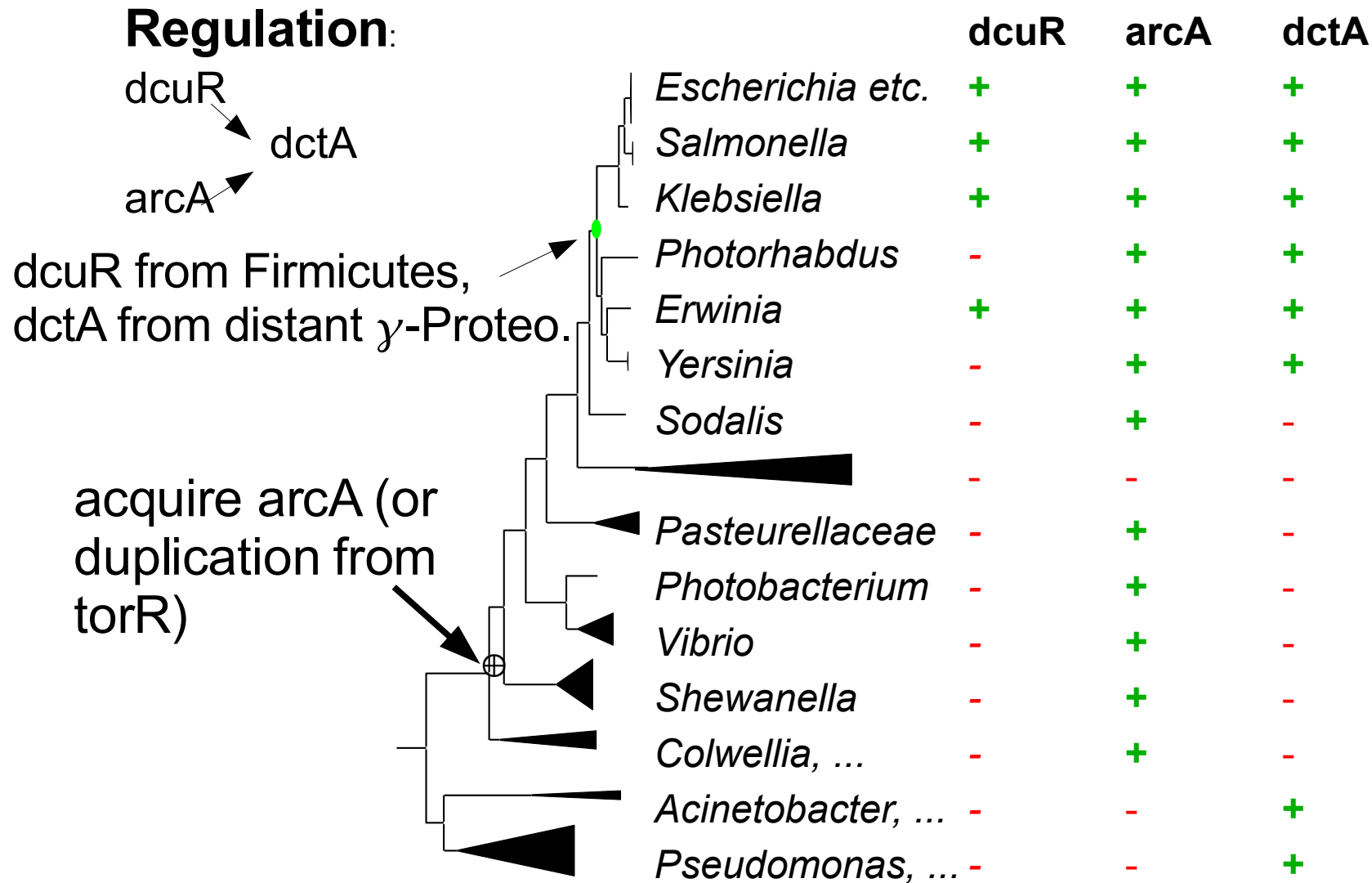
(A) Niche-Specific Neighbor Regulators vs. Conserved Global Regulators

Tree of Life

Co-transfer of regulator with operon

Conserved Global Regulator

HGT

Neighbor | Operon

GR

Last ~1 Billion Years

and gain a new site

Neighbor | Operon

GR

Repeated transfers between related bacteria

Time

and transfer again with the site

HGT

Neighbor | Operon

GR
(not HGT)

(B) Rapid & Convergent Evolution of Regulatory Interactions

1. Duplication & conservation
(5%-8%)

2. HGT & convergence
(5%-6%)

3. Duplication & "convergence"
(1%-3%)

4. Other (85%)

Time

TF → Operon

TF1 → Operon
TF2 → Operon

Acquire a paralog

HGT

TF1 → Operon
TF2 ✗

Evolve a new site
(not shared)

TF1 → Operon
TF2 →

Acquire an operon

TF1 ✗ → Operon
TF2 ✗

Evolve a shared site
(paralogous TFs with similar DNA specificity)

TF1 → Operon
TF2 →

HGT or duplication, then divergence

TF1 → Operon
TF2 ✗
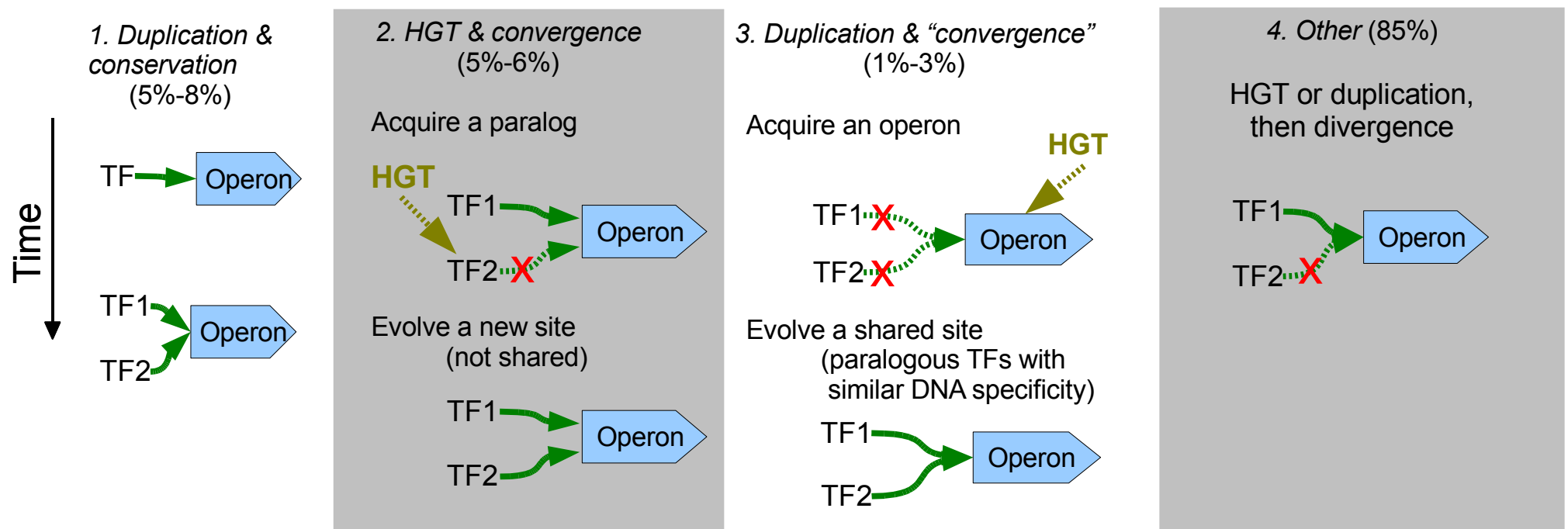
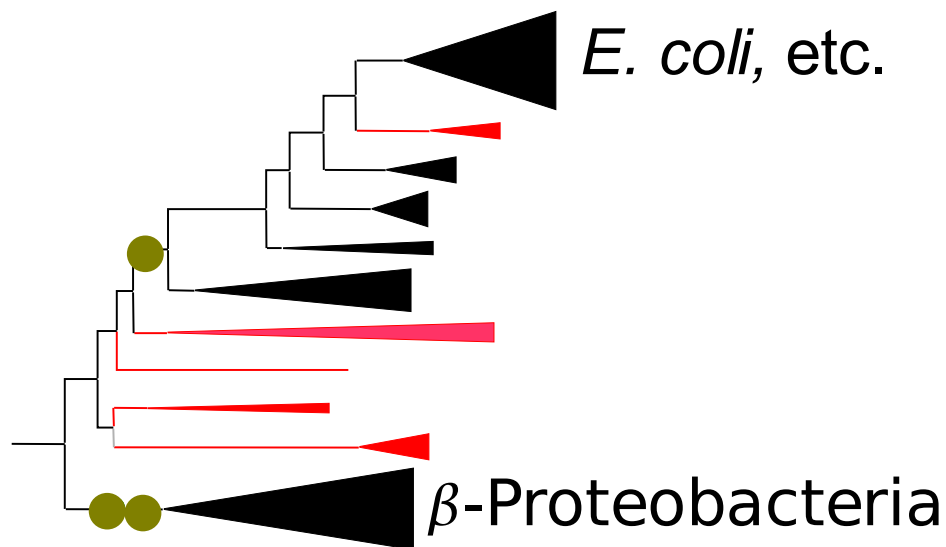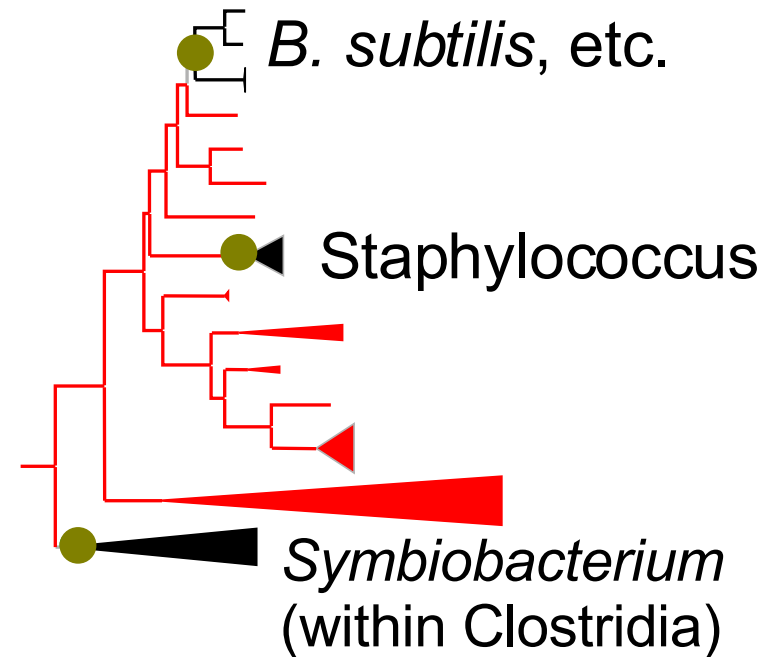# BBHs of TFs Have Different Functions

- Different pathways & stimuli

  - *E. coli* betI: choline -> osmotic stress

  - *B. subtilis* pksA: polyketide synthase

betI: HGT or 3 losses

E. coli, etc.

β-Proteobacteria

pksA: 2 HGT or >5 losses

B. subtilis, etc.

Staphylococcus

Symbiobacterium (within Clostridia)

17/26 different functions between *E. coli, B. subtilis*
9/20 different between *E. coli* and other divisions of Proteobacteria